A Multi-Modal Decision Support System with Allergy-Aware Recipe Understanding Powered by a Plan Representation

Vedant Khandelwal[1], Kausik Lakkaraju[1], Vishal Pallagani[1], Priyadharsini Ramamurthy[2], Revathy Venkataramanan[1], Sathyanarayanan N. Aakur[2], Biplav Srivastava[1]

[1]AI Institute, University of South Carolina
[2]Oklahoma State University
{vishalp@mailbox.,vedant@mailbox.,revathy@email.,kausik@email.,biplav.s@}sc.edu
{pramamu@ostatemail.,saakurn@}okstate.edu

Abstract

Food is not only a basic human necessity, but also a key factor driving a society's health and economic well-being. The need to build and deploy decision-support tools (AI) in the food domain has been on the rise. However, AI must understand the concepts in the food domain (e.g., recipes, ingredients), be tolerant to failures encountered while cooking (e.g., browning of butter), handle allergy-based substitutions, and inter-relationships between multiple data modalities. We introduce a rich recipe representation (R3), which is inspired by the plan representation of PDDL. R3 consists of atomic actions with well defined semantics, spanning multiple modalities, and enriched with external knowledge to enable building of robust AI solutions. We demonstrate how R3 can be used in building multi-modal decision support system that is capable of performing constrained queries, with a special focus on allergens. See our demo at https://youtu.be/xldeePzS7Ko.

Introduction

Maintaining a healthy diet is essential for good health and plays a vital role in managing chronic diseases. The importance of diet and food has attracted increasing attention leading to *precision nutrition* (Chatelan, Bochud, and Frohlich 2019), an area of research which aims to personalize the dietary intake based on unique characteristics of an individual. The process of cooking is a key factor for precision nutrition, which has led to extensive research in building and deploying decision-support artificial intelligence (AI) tools - varying from information retrieval (IR) interfaces to task-oriented chatbots.

In-spite of advanced computing approaches and AI, a persistent challenge is to glean meaningful insights from big data. It is essential for the AI models to understand and reason about the concepts in food domain. In contrast, the recipes are made available as textual documents which makes it difficult for machines to read, reason and handle ambiguity. This demands a need for better representation for the recipes in order to facilitate the development of robust decision support systems (DSS). On the other hand, the planning community has seen many successful representations which have helped navigate complex domains. Taking cues from this, we adopt the plan representation of PDDL for recipes but do not assume how recipes may have been created. We thus allow not only efficient management of recipe data (storage and retrieval) but also enable future automation in recipe creation (by reuse and composition of recipes to create new recipes), and food preparation (by monitoring the execution of agents) by allowing reasoning with both the data and control flow inherent in the plan-based recipe representation consisting of text and images. The following sections give a brief description of the decision support system here; more details about R3 and the overall functioning of the system can be found in (Pallagani et al. 2022)

System Description

We explain the planning-inspired modeling of R3 from natural language cooking instructions, and demonstrate its capabilities in performing constrained information retrieval (CIR). Figure 1 shows the overall architecture of the proposed system, which shows the construction of R3 and its role in performing the constrained multi-modal queries obtained from user.

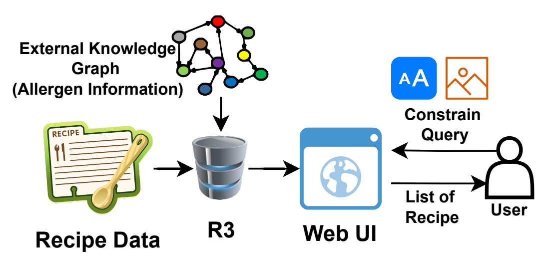


Figure 1: System Architecture

R3 - Rich Recipe Representation

We define **R3**, as a tuple consisting of the metadata harvested from the original recipe dataset (Yagcioglu et al. 2018), alongwith additional information such as planning-inspired predicates for cooking instructions, allergens, and multi-modal data points. Figure 2 captures an

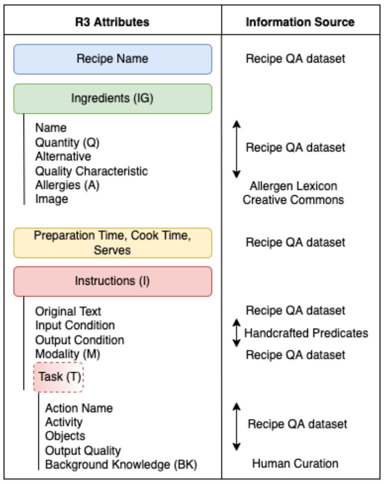
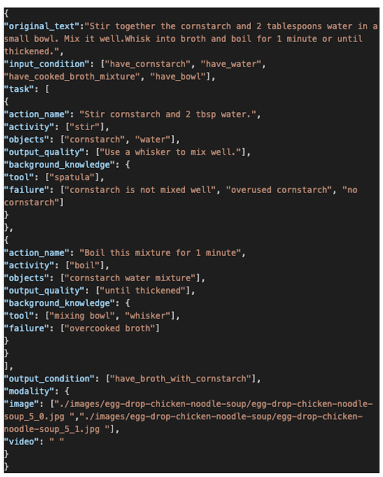


Figure 2: Information source for different attributes in **R3**.

overall view of which source is used to populate R3. Figure 3, shows an example of an instruction in R3.



```
{
"original_text":"Stir together the cornstarch and 2 tablespoons water in a
small bowl. Mix it well.Whisk into broth and boil for 1 minute or until
thickened.",
"input_condition": ["have_cornstarch", "have_water",
"have_cooked_broth_mixture", "have_bowl"],
"task": [
{
"action_name": "Stir cornstarch and 2 tbsp water.",
"activity": ["stir"],
"objects": ["cornstarch", "water"],
"output_quality": ["Use a whisker to mix well."],
"background_knowledge": {
"tool": ["spatula"],
"failure": ["cornstarch is not mixed well", "overused cornstarch", "no
cornstarch"]
}
},
{
"action_name": "Boil this mixture for 1 minute",
"activity": ["boil"],
"objects": ["cornstarch water mixture"],
"output_quality": ["until thickened"],
"background_knowledge": {
"tool": ["mixing bowl", "whisker"],
"failure": ["overcooked broth"]
}
}
],
"output_condition": ["have_broth_with_cornstarch"],
"modality": {
"image": ["./images/egg-drop-chicken-noodle-soup/egg-drop-chicken-noodle-
soup_5_0.jpg ","./images/egg-drop-chicken-noodle-soup/egg-drop-chicken-
noodle-soup_5_1.jpg "],
"video": " "
}
}
```


Figure 3: Example of an instruction in R3 (Original Text: Stir together the cornstarch and 2 tbsps water in a small bowl. Mix it well. Whisk into broth and boil for 1 minute or unitl thickened.)

Constrained Information Retrieval

CIR is concerned with obtaining specific information from the given data based on the constraints mentioned by the user. CIR system currently can handle the following types of constrained queries from the user:

- **Outcome:** The queries that impose constraints on the outcome of the recipe, i.e., filtering recipe by ingredient name/allergen fall under the category of outcome constraints.
- **Process:** The queries that impose constraints on the process of the recipe, i.e., filtering recipes by length/type of cooking (e.g., boiled, deep-fried) contribute to process constraints.

The similarity metric employed to match the user query with attributes in R3 is Levenshtein distance. For image as an input, which can be of an ingredient or final cooked recipe, similarity is computed using the Scale Invariant Feature Transform (SIFT) algorithm.

Demonstration

The user is equipped with a web-based interface, as shown in Figure, to perform CIR on the rich representation for the recipes. The system interface offers the functionality of using *text*, *images* or a *combination of both* to perform constrained queries. We perform an evaluation of the system using 50 unique queries based on 25 unique enriched recipe, consisting of both process and outcome constraints. The queries are generated randomly to avoid the risk of cherry-picking, and allows us to test our system on various edge cases. In Figure 4, the user asks for a recipe containing bacon by uploading an image of bacon to the interface. Here, out of 25 unique recipes, the system shows top six filtered recipe and their images based on the query constraint provided. The filtered set of recipes are displayed in rank order based on the percentage similarity. The details about allergen and a link to the json object of the filtered recipe is also provided. Recipe representation can help

Figure 4: Result of user query of asking recipes containing bacon

answer a wider set of queries, out of which the queries related to allergens are of utmost importance. The proposed system is able to handle queries with both implicit and explicit allergen information based queries, where allergen category name is provided and non provided to the system explicitly, respectively.

References

Chatelan, A.; Bochud, M.; and Frohlich, K. L. 2019. Precision nutrition: hype or hope for public health interventions to reduce obesity? *International journal of epidemiology*, 48(2): 332–342.

Pallagani, V.; Ramamurthy, P.; Khandelwal, V.; Venkataramanan, R.; Lakkaraju, K.; Aakur, S. N.; and Srivastava, B. 2022. A Rich Recipe Representation as Plan to Support Expressive Multi Modal Queries on Recipe Content and Preparation Process.

Yagcioglu, S.; Erdem, A.; Erdem, E.; and Ikizler-Cinbis, N. 2018. RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes. arXiv:1809.00812.